Title
Multiple interactive memory representations underlie the induction of false memory.

Journal
Proceedings of the National Academy of Sciences of the United States of America, 116(9)

ISSN
0027-8424

Authors
Zhu, Bi
Chen, Chuansheng
Shao, Xuhao
et al.

Publication Date
2019-02-01

DOI
10.1073/pnas.1817925116

Peer reviewed

Multiple interactive memory representations underlie the induction of false memory

Bi Zhu[a,b,c,d], Chuansheng Chen[e], Xuhao Shao[a,b], Wenzhi Liu[a,b], Zhifang Ye[a,b], Liping Zhuang[a,b], Li Zheng[a,b], Elizabeth F. Loftus[e,1], and Gui Xue[a,b,1]

[a]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China; [b]IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China; [c]Institute of Developmental Psychology, Beijing Normal University, Beijing 100875, China; [d]Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing 100875, China; and [e]Department of Psychological Science, University of California, Irvine, CA 92697

Contributed by Elizabeth F. Loftus, January 15, 2019 (sent for review October 18, 2018; reviewed by Nancy Dennis and Steve Ramirez)

Theoretical and computational models such as transfer-appropriate processing (TAP) and global matching models have emphasized the encoding–retrieval interaction of memory representations in generating false memories, but relevant neural mechanisms are still poorly understood. By manipulating the sensory modalities (visual and auditory) at different processing stages (learning and test) in the Deese–Roediger–McDermott task, we found that the auditory-learning visual-test (AV) group produced more false memories (59%) than the other three groups (42∼44%) [i.e., visual learning visual test (VV), auditory learning auditory test (AA), and visual learning auditory test (VA)]. Functional imaging results showed that the AV group's proneness to false memories was associated with (*i*) reduced representational match between the tested item and all studied items in the visual cortex, (*ii*) weakened prefrontal monitoring process due to the reliance on frontal memory signal for both targets and lures, and (*iii*) enhanced neural similarity for semantically related words in the temporal pole as a result of auditory learning. These results are consistent with the predictions based on the TAP and global matching models and highlight the complex interactions of representations during encoding and retrieval in distributed brain regions that contribute to false memories.

false memory | study modality | fMRI | visual | auditory

Memory is the ability to encode, store, and retrieve information that we encounter in our environment. However, it is not always reliable. Ample evidence has shown that both exogenous misinformation (1) and endogenous memory distortion (2) can lead to false memories. The prevalence of false memories is also affected by other factors such as the sensory modality in which the information is initially learned and subsequently tested (3). For example, in the Deese–Roediger–McDermott (DRM) task, the auditory-study visual-test (AV) condition has been shown to produce more false memories compared with the visual-study visual-test (VV) condition (3–5). Perhaps not coincidentally, most previous misinformation studies that created high rates of false memories also chose to plant false information in a one particular sensory modality (e.g., verbally retelling a story) which was different from the modality in which the original information was acquired (e.g., viewing a scene) (6–8).

Theoretical and computational models have long emphasized the interactions of memory representations during encoding and retrieval in determining both true and false memories (9). For example, according to the transfer-appropriate processing (TAP) (10) and encoding specificity (11) hypotheses, successful memory performance tends to occur when there is a substantial similarity between the encoding and retrieval processes. The TAP hypothesis gains support from a vast body of behavioral research (12). For example, better memory can be achieved when the modality or context of initial study and that of subsequent testing is matched (13). Such a match of context has been found to increase true memories and reduce false memories (3). Using multivoxel pattern analysis (MVPA) that could examine the representational content carried by the neural activation pattern, studies have tested the TAP model directly. The results indicated that the match in context during encoding and retrieval enhances the neural pattern similarity between encoding and retrieval (i.e., encoding–retrieval similarity), which in turn increases the strength of true memory (14).

In addition to the encoding–retrieval interaction of representations for the same item, emerging evidence suggests that the memory of a given item also depends on its interaction with all other items in the same episodic space as well as the knowledge stored in long-term memory (9). As an extension of the TAP model, the global matching hypothesis (15, 16) posits that the memory strength of a given item derives from the match (measured as similarity) between its representation and those of all other studied items in the episodic memory space (17–20). Moreover, false memory in the DRM paradigm is a result of many partial matches between the representations of critical lures and those of studied items, leading to strong overall matches with memory traces (2). By integrating the TAP and global matching mechanisms, a previous MVPA study calculated the encoding–retrieval neural global pattern similarity (ER-nGPS), which reflected the

Significance

False memories appear in our daily life due to the reconstructive nature of memory. They are affected by the contexts of both learning and testing. The combination of auditory learning and visual test (AV) resulted in more false memories compared with other three combinations of sensory modalities during learning and test (VV, VA, and AA). Using sophisticated neural representation analysis of fMRI data, we found that this effect was jointly related to three neural mechanisms: Compared with VV, AV showed weaker memory signals in the visual cortex, reduced prefrontal monitoring, and a greater reliance on semantic encoding during learning. These mechanisms highlight the complex interactions of memory representations during encoding and retrieval that give rise to the appearance of false memories.

Author contributions: B.Z., C.C., E.F.L., and G.X. designed research; B.Z., X.S., W.L., L. Zhuang, and L. Zheng performed research; B.Z., X.S., Z.Y., and G.X. analyzed data; and B.Z., C.C., E.F.L., and G.X. wrote the paper.

Reviewers: N.D., Pennsylvania State University; and S.R., Boston University.

The authors declare no conflict of interest.

Data deposition: fMRI data and materials are available at https://openneuro.org/datasets/ds001650.

[1]To whom correspondence may be addressed. Email: eloftus@uci.edu or gxue@bnu.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1817925116/-/DCSupplemental.

Published online February 14, 2019.

representational match between a given item during retrieval and all other studied items during encoding (21). They found that the ER-nGPS in distributed brain regions could predict the strength of both true and false memories. Besides the representation in the episodic memory space, the semantic similarity in the temporal pole (i.e., schema), which reflects one's semantic organizations, could also predict both true and false memories, reflecting the adaptive nature of false memories (22).

Importantly, studies have further revealed that multiple representations and processes may contribute differentially to true and false memories. For example, a recent meta-analysis of studies using univariate analysis suggests that the frontal cortex shows greater activations for false memory than true memory, whereas the visual cortex shows mixed results varying by the type of baselines and modality of stimulus (23). The MVPA study suggests that whereas the ER-nGPS in the frontoparietal cortex is associated with both true and false memories, the ER-nGPS in the visual cortex is only associated with true memory for visually presented learning materials (21). Interestingly, the increased activation in the prefrontal cortex for lures, compared with foils and targets, was mainly due to the fact that lures (but not foils and targets) involved discrepant memory signals between the visual cortex and the frontoparietal cortex (21). Nevertheless, no study has thus far examined how the contextual match may modulate these mechanisms to influence the rate of false as well as true memories.

The current study addressed these issues with two experiments. In Exp. 1, we used the DRM paradigm to examine the modality effect by comparing four experimental conditions in a behavioral study: the AV, VV, auditory-learning auditory-test (AA), and visual-learning auditory-test (VA) conditions. Based on the behavioral results of Exp. 1, Exp. 2 compared the VV and AV conditions with fMRI data to examine the neural mechanisms associated with elevated false memories under the AV condition. Using fMRI and MVPA, we tested three main hypotheses derived from the global matching model and TAP hypothesis. First, the ER-nGPS in the frontoparietal cortex should be comparable for AV and VV conditions, because this brain region contains supramodal representation supporting both true and false memories. Second, the VV condition would show greater ER-nGPS in the visual cortex than would the AV condition, because the latter involved modality mismatch. Furthermore, we predicted that ER-nGPS in the visual cortex would contribute to true memory for the VV condition but not the AV condition and would not contribute to false memories in either condition. Third, we predicted that as no memory signal in the visual cortex was available to differentiate true and false memories in the AV condition, the prefrontal cortex would show a smaller increase for lures relative to targets and foils in the AV than in the VV condition, resulting in more false memories. Furthermore, since previous studies suggested that semantic similarity and neural semantic representation in the temporal pole could predict false memories (22), we further examined which modality could modulate the involvement of semantic coding and false memories. Our results provide important neural evidence to highlight the complex interactions of memory representations during encoding and retrieval in creating false memories as well as true ones (9).

Results

Auditory Encoding and Visual Retrieval (AV) Elicited the Highest Rate of False Memories. Given the critical role of experimental parameters in determining the prevalence of false memories, we first conducted a behavioral study (Exp. 1) to examine the modality effect on false recognition, using a slow DRM paradigm that would be needed in the subsequent fMRI study to estimate the blood oxygen level-dependent signal pattern associated with each trial (*Methods* and Fig. 1*A*). Participants made pleasantness judgment on nine word lists, each containing eight semantically related words. After a 10-min filler task, participants took a recognition test consisting of 36 targets (studied), 36 lures (semantically related but unstudied), and 36 foils (unrelated and unstudied). A total of 118 healthy young college students were randomly assigned into one of the four experimental groups (i.e., AV, VV, VA, and AA). All experimental paradigms and parameters (including the stimulus duration) were matched across the groups except for the sensory modality of the stimuli during learning and test.

Results showed substantial rates of false memories in each of the four groups. During the test, between 42% and 59% of the critical lures were judged as old items (LO), compared with between 5% and 16% of the foils judged as old (FO) ($P < 0.001$) (*SI Appendix*, Table S1). Critically, the false-memory rates were significantly higher in the AV group (59%) than in the other three groups (42 ~ 44%) ($P < 0.002$), the latter of which did not differ among themselves ($P > 0.20$). Using the corrected false-memory rate (i.e., LO − FO), which controlled for guessing, we still found that the rates of false memories in the AV group (43%) were significantly greater than in the VV group (36%) ($P = 0.04$), and marginally greater than the VA group (37%) ($P = 0.08$). No other group differences were significant ($P > 0.25$) (Fig. 1*B* and *SI Appendix*, Table S1). It is worth noting that corrected true memory [i.e., TO (targets judged as old) − FO] was lower in the AV group (73%) than the other three groups (81, 83, and 81% for VV, AA, and VA, respectively) ($P < 0.02$), with no differences among the latter three groups ($P > 0.50$). Simply put, the results suggest that the AV condition induced significantly higher rates of false memory and lower rates of true memory than the VV condition. The reaction time data indicated that subjects were faster to recognize studied items and reject unrelated foils than to recognize and reject lures, and slower to make judgments in auditory presentations than in visual presentations (*SI Appendix*, Fig. S1*A*).

Memory Under Auditory Encoding Relied More on Semantic Information than on Visual Encoding. A previous study suggests that semantic similarity across items has a greater effect on false memories than on true memories (21). In an exploratory analysis, we examined whether stimulus modality would modulate this semantic similarity effect. To do this, we asked subjects to rate the semantic similarity among all words (targets and lures) within each list after the memory test and then calculated the semantic global similarity (sGS) for each item (target or lure) by averaging a given item's semantic similarity with all other studied items belonging to the same list. For lures, there were positive correlations between the sGS and the memory of lures in all four groups [r (34) = 0.46, 0.50, 0.47, and 0.55, $P < 0.005$, for AV, VV, AA, and VA, respectively] (*SI Appendix*, Fig. S2*A*). For targets, there were also positive correlations between the sGS and the memory of targets in the AV [r (34) = 0.40, $P = 0.02$] and AA groups [r (34) = 0.41 $P = 0.01$], but not in the VV and VA groups [r (34) = 0.17 and −0.09, $P > 0.31$, for the VV and VA groups, respectively]. Direct comparison of these correlations suggested that semantic similarity produces stronger effects (i.e., more positive correlations) on targets under the two auditory study conditions (i.e., AA and AV) than the VA condition ($P = 0.03$ and 0.04), but not the VV condition ($P = 0.28$ and 0.31). It should be noted that these differences could not be accounted for by the differences in the range (or SDs) of their distributions (*SI Appendix*, Table S2).

Replication of the Behavioral Results in an Independent Sample of the fMRI Study. To reiterate, the behavioral experiment revealed higher rates of false memories in the AV than in the other three conditions. To examine the neural mechanisms, we then did an fMRI study (i.e., Exp. 2) by recruiting an independent sample of

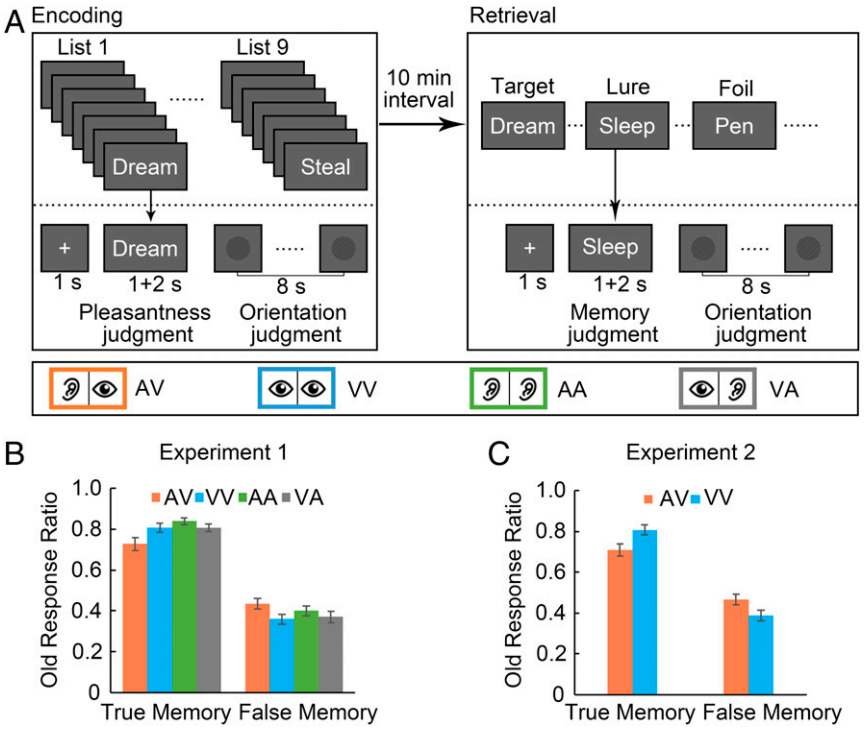

Fig. 1. Experimental procedure and behavioral results of modality effect on true and false memories. (*A*) Experimental procedure in AV, VV, AA, and VA groups. At encoding, 72 words (nine lists of eight semantically related words) were presented visually for VV and VA and auditorily for AV and AA. Participants were asked to make a pleasantness judgment on each word. At retrieval, 108 words (36 targets, 36 lures, and 36 foils) were presented visually for VV and AV and auditorily for AA and VA. Participants were asked to make a memory judgment on each word. The slow event-related design (each trial lasting 12 s) was used for both encoding and retrieval phases. Each word was presented for 1 s, with another 2 s for judgment. (*B* and *C*) The corrected true memory (i.e., the endorsement rate of targets judged as old minus that of foils judged as old) and corrected false memory (i.e., the endorsement rate of lures judged as old minus that of foils judged as old) in Exp. 1 (*B*) and Exp. 2 (*C*). Error bars indicate between-participant SEs. AV had higher false memory and lower true memory than did VV in both experiments.

59 participants and randomly assigning them into the AV (n = 30; 15 females) and VV (n = 29; 13 females) groups. We chose the VV condition as the reference group because VV and AV were the most frequently studied conditions in previous behavioral and fMRI studies of false memory (23), and the representational mechanisms under the VV condition have been extensively examined previously (21). Behavioral results of Exp. 2 replicated those of Exp. 1 (*SI Appendix*, Table S1). Specifically, there were more false memories of the critical lures ($P = 0.006$) and fewer true memories ($P = 0.02$) for AV than for VV, but no significant difference in false reporting of foils ($P = 0.21$). Again, after correcting for guessing, we still found more false memories in AV (47%) than in VV (39%) ($P = 0.04$), and fewer corrected true memories in AV (71%) than in VV (81%) ($P = 0.02$) (Fig. 1*C* and *SI Appendix*, Table S1). The reaction time data revealed that subjects were faster in VV than in AV ($P < 0.001$), but there was no significant modality-by-response interaction [$F(3,171)$ = 1.86, $P = 0.14$] (*SI Appendix*, Fig. S1*B*). Finally, we again found that semantic similarity had similar effects (i.e., positive correlations) on lures under both AV [r (34) = 0.51, $P = 0.001$] and VV [r (34) = 0.50, $P = 0.002$] conditions (difference in correlations, $P = 0.93$) but stronger effects (i.e., more positive correlations) on targets under the AV condition [r (34) = 0.57, P = 0.0003] than under the VV condition [r (34) = 0.09, $P = 0.62$] (difference in correlations, $P = 0.02$) (*SI Appendix*, Fig. S2*B*), suggesting that auditory encoding led to greater semantic associations than did visual encoding.

ER-nGPS in the Left Middle Frontal Gyrus Predicted both True and False Memories in VV and AV. Having demonstrated that AV elicited more false memories than did VV, we then compared the neural representations of true and false memories between these two conditions. In particular, we integrated both the TAP and global matching mechanisms by calculating the ER-nGPS, which reflected the neural activation pattern similarity (Pearson r) of each test item at retrieval with the neural activation pattern of all 72 studied items at encoding (Fig. 2*A*). We predicted that ER-nGPS in the frontoparietal region should be positively associated with memory strength for both true and false memories, under both AV and VV conditions, but the ER-nGPS in the visual cortex should be specific to the VV condition due to the modality mismatch under the AV condition.

For the VV condition, a whole-brain searchlight analysis revealed that the ER-nGPS associated with true memory strength [i.e., TO vs. FN (foils judged as new)] was located in the left middle frontal gyrus [LMFG; Montreal Neurological Institute (MNI): −50, 16, 30, Z = 5.83] and the left dorsal lateral occipital complex (LdLOC; MNI: −30, −64, 38, Z = 5.30). For the AV condition, the ER-nGPS associated with true memory strength (i.e., TO vs. FN) was found in the LMFG (MNI: −34, 18, 24, Z = 4.43), the right inferior frontal gyrus (RIFG; MNI: 46, 6, 18, Z = 3.74), the paracingulate gyrus (PACG; MNI: −6, 30, 28, Z = 3.99), the left supramarginal gyrus (LSMG; MNI: −64, −38, 32, Z = 3.81), the left superior parietal lobule (LSPL; MNI: −30, −56, 44, Z = 4.57), the lingual gyrus (LING; MNI: −8, −60, 2, Z = 3.95), and the right precuneus (PCUN; MNI: 12, −62, 24, Z = 3.49). Conjunction analysis revealed two overlapping clusters across AV and VV conditions, the LMFG (MNI: −54, 10, 38, Z = 3.87) and LING (MNI: −8, −60, 2, Z = 3.80).

Focusing on these regions, we further examined the hypothesis that the ER-nGPS in these regions were also associated with

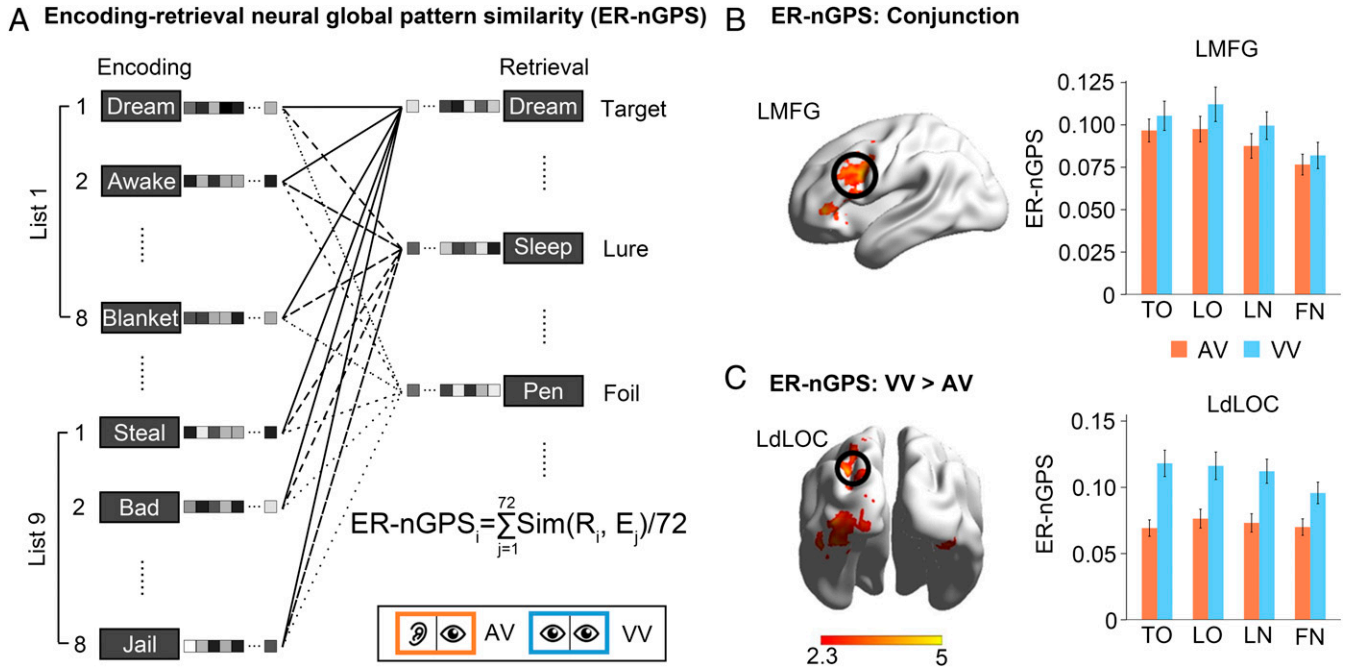

Fig. 2. The ER-nGPS in the VV and AV groups. (*A*) The ER-nGPS was calculated by averaging the Fisher's Z scores reflecting neural activation pattern similarity of each test item with the neural activation patterns of all 72 studied items. It was calculated for the VV and AV groups, separately. (*B*) Conjunction analysis for ER-nGPS in VV and AV. The ER-nGPS in the LMFG was associated with both true and false memories (i.e., TO > FN and LO > LN) for both VV and AV. (*C*) ER-nGPS VV > AV: The ER-nGPS in the visual cortex (e.g., LdLOC) was associated with greater true memory (i.e., TO > FN) in VV than AV. In these brain images, results were rendered onto a population-averaged atlas using the software Connectome Workbench (61). The bar graphs of ER-nGPS, as a function of memory status, are shown for brain ROIs. Error bars indicate between-participant SEs.

false memories. Repeated-measures ANOVA revealed a significant main effect of false memory [LO vs. LN (lures judged as new)] in the LMFG [$F(1,57) = 9.82$, $P = 0.003$] but not in the LING [$F(1,57) = 0.90$, $P = 0.35$]. No effect of experimental condition or condition-by-false-memory interaction was found in these regions (Fig. 2*B*). Together, our results suggested that the ER-nGPS in the LMFG was predictive of both true and false memories, whereas the ER-nGPS in the LING was predictive of true memory, regardless of sensory modality of learning.

The ER-nGPS in the Visual Cortex Predicted True Memory in VV but Not in AV. Having identified the common mechanisms for VV and AV conditions, we further examined whether the VV condition had any additional true memory signal (i.e., TO − FN) compared with the AV condition. We did not directly compare between AV and VV as it could be confounded by general task differences due to modality mismatch. As predicted, a whole-brain direct comparison (i.e., VV [TO-FN] − AV [TO-FN]) revealed greater ER-nGPS for VV than AV in the bilateral visual cortex, including the LdLOC (MNI: −32, −66, 38, Z = 4.27), the left ventral lateral occipital complex (LvLOC; MNI: −42, −92, −2, Z = 4.48), and the right ventral lateral occipital complex (RvLOC; MNI: 38, −76, −16, Z = 4.25). No brain region showed greater ER-nGPS for TO-FN in AV than VV. Post hoc paired-sample tests in the above regions revealed that in the LdLOC the ER-nGPS was significantly associated with true memory strength (i.e., TO > FN) in VV ($P < 0.001$), but not in AV ($P = 0.70$) (Fig. 2*C*). A similar pattern was found in the LvLOC and RvLOC (*SI Appendix*, Fig. S3).

Also consistent with our hypothesis, there was no significant interaction between modality (AV vs. VV) and false memory (LO vs. LN) in these visual regions of interest (ROIs). A further analysis revealed that these regions were not associated with false memory strength, that is, no significant differences between LO and LN for either VV ($P > 0.46$) or AV ($P > 0.15$). Together, these results suggest that the ER-nGPS in the visual cortex was associated with true memory, and it was stronger for VV than for AV.

The ER-nGPS in the Left Hippocampus Predicted True Memory in VV but Not in AV. The whole-brain searchlight analysis revealed no effect of ER-nGPS in the hippocampus. The hippocampus is a crucial area for memory (24), particularly in context representation (25), and previous studies have suggested that the global matching signal in the hippocampus was associated with subsequent true memory (26, 27). We then examined the effects of modality and memory type on the ER-nGPS in the left and right hippocampus (LHIP and RHIP), using anatomically defined ROIs according to the Harvard–Oxford template (*SI Appendix*, Fig. S4). We found greater ER-nGPS for VV than AV in both LHIP [$F(1,57) = 4.31$, $P = 0.03$] and RHIP [$F(1,57) = 6.82$, $P = 0.01$]. Importantly, there was a significant interaction between modality (AV vs. VV) and memory type (TO vs. FN) in the LHIP ($P = 0.04$), though not in the RHIP ($P = 0.22$). In the LHIP, the ER-nGPS was higher for TO than FN in VV ($P = 0.004$) but not in AV ($P = 0.60$), suggesting that the ER-nGPS in the LHIP was associated with true memory strength, and, consistent with its role in context representation, ER-nGPS was stronger when the contexts of encoding and retrieval were matched.

AV Showed Weaker Prefrontal Monitoring During Retrieval than Did VV. The above analyses confirmed our first two hypotheses that (*i*) VV and AV would have comparable memory signals supporting both true and false memories in the prefrontal cortex and (*ii*) AV would have a weaker true memory signal in the visual cortex than would VV. As a result, the lures in the VV condition could lead to a mismatch of memory signals (as indexed by the ER-nGPS) between the prefrontal cortex (for both targets and lures) and the visual cortex (for targets only), which triggers

prefrontal monitoring processes at retrieval that could reject lures and reduce false memory (21, 28, 29). In contrast, such a mechanism should be weakened in the AV condition as both targets and lures rely solely on the frontal memory signals, and no memory signal for targets in the visual cortex could help to differentiate true and false memories. Consistently, although both AV and VV showed greater frontal activations to lures compared with foils judged as new (*SI Appendix*, Tables S3 and S4), direct comparison (i.e., VV [Lure-FN] − AV [Lure-FN]) revealed greater activation in the left lateral prefrontal cortex (LPFC; MNI: −42, 28, 22, Z = 3.57) (Fig. 3 *A* and *B*). No brain region showed greater univariate activations for Lure-FN in AV than in VV.

If the increased LPFC activation truly reflected the discrepancy between the different memory signals, it is reasonable to predict that the LPFC activation at retrieval would be positively correlated with the ER-nGPS mismatch for lures between the LMFG (which reflected both true and false memories) and LvLOC (which only reflected true memory). Indeed, we found such positive correlations for lures in the VV group [r (27) = 0.62, $P < 0.001$] as well as in the AV group [r (28) = 0.43, P = 0.02] (Fig. 3 *C* and *D*).

AV Showed Greater Encoding Neural Global Semantic Similarity in the Temporal Pole than Did VV. The above analyses confirmed our main hypotheses that, compared with the VV condition, the AV condition would lead to lower true memory signals in the visual cortex and weakened monitoring process in the prefrontal cortex. However, if modality mismatch was the only cause of elevated false memories for AV, we would have expected comparable false memories for AV and VA conditions, which was consistently not the case. So what was special about auditory learning? Our behavioral analysis showed that semantic similarity had a greater effect on memory under the two auditory learning conditions than under the two visual learning conditions, suggesting greater semantic coding for the former. A previous study also suggests that the semantic organization in the temporal pole could also predict both true and false memories (22). As a result, one would predict greater neural similarity across encoded items for AV than VV in the brain regions representing semantic information, such as the temporal pole. To test this hypothesis, we computed the neural global semantic similarity during encoding (EnGSS) for trials belonging to the same semantic list (Fig. 4*A*). In particular, the EnGSS was calculated by averaging the pairwise correlations of the neural activation pattern of all eight words within a list and then averaging them across all nine lists. If there was greater semantic encoding for AV than VV, there should be greater EnGSS for AV than VV in regions involved in semantic representations. Consistent with our hypothesis, a whole-brain searchlight analysis showed greater EnGSS for AV than VV in the left planum temporale (LPT; MNI: −62, −14, 6, Z = 8.47) that extended to the left temporal pole (LTP; MNI: −60, 8, −10, Z = 4.99), the right central opercular cortex (ROC; MNI: 62, −14, 12, Z = 6.68) that extended to the right temporal pole (RTP; MNI: 60, 6, 0, Z = 5.78), and the left occipital fusiform gyrus (LOF; MNI: −22, −76, −22, Z = 5.55) (Fig. 4*B*). No brain region showed higher EnGSS for VV than AV.

Focusing on the LTP and RTP, we further explored whether their EnGSS could predict false memories. We found that across subjects and conditions there was a significant positive correlation between EnGSS and false memory rate in the RTP [r (57) = 0.35, P = 0.007], and a numerical trend in the LTP [r (57) = 0.20, P = 0.12] but no correlation with true memory rate [r (57) = −0.05 and −0.11, $P > 0.41$] (Fig. 4*C*). Moreover, across the nine lists of words, we found a significant positive correlation between EnGSS in the LTP and the mean false memory rate (averaged across four critical lures for each word list) for the AV group [r (7) = 0.72, P = 0.03], but not for the VV group [r (7) = −0.46, P = 0.22]. No such correlation was found in the RTP for either AV or VV group [r (7) = 0.47 and −0.28, $P > 0.20$]. These results suggested that the semantic representation in the temporal pole predicted higher false memories under the AV than VV condition.

Controlling the Differences in the Univariate Activation Level at Retrieval. The above analysis revealed significant differences in neural pattern similarity between AV and VV, which could account for the former's elevated false memories. To make sure that these differences were not caused by univariate activation levels, we did an additional univariate analysis. The results revealed no significant difference between the VV and AV conditions for either true memory (TO vs. FN) or false memory (LO vs. LN). Conjunction analyses revealed the regions showing common effects for both VV and AV conditions. Of these regions, we found greater activations for TO than FN (i.e., true memory) in the left frontal pole, left superior frontal gyrus, PCUN, left dorsal occipital cortex, and right occipital pole (*SI Appendix*, Fig. S5 and Table S4; also see *SI Appendix*, Figs. S6 and S7 and Table S3 for results for AV and VV, separately) but no significant difference for false memory (i.e., LO vs. LN).

We further conducted mixed-effect regression analyses to control for the effect of univariate activation levels on the ER-nGPS. These results suggested that, after controlling for the univariate activation level, the ER-nGPS in the LMFG was still associated with true and false memories ($P < 0.003$) and the ER-nGPS in the LING was still associated with true memory ($P < 0.001$), but not associated with false memory (P = 0.37). Similarly, after controlling for the univariate activation levels, the interactions between modality (AV vs. VV) and true memory

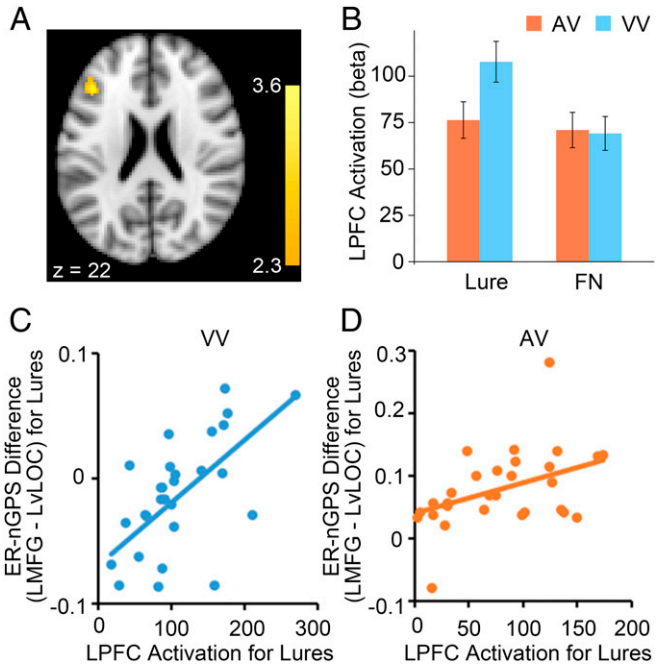

Fig. 3. Greater prefrontal activation for monitoring process at retrieval in the VV group than in the AV group. (*A*) The LPFC showed higher levels of univariate brain activation for lures than foils judged as new (FN) at retrieval in VV than in AV. (*B*) Bar graph of the retrieval neural activation level in the LPFC as a function of memory status (Lure vs. FN). Error bars indicate between-participant SEs. (*C* and *D*) Greater LPFC univariate brain activations for lures were associated with higher ER-nGPS difference (LMFG − LvLOC) for lures in VV and AV.

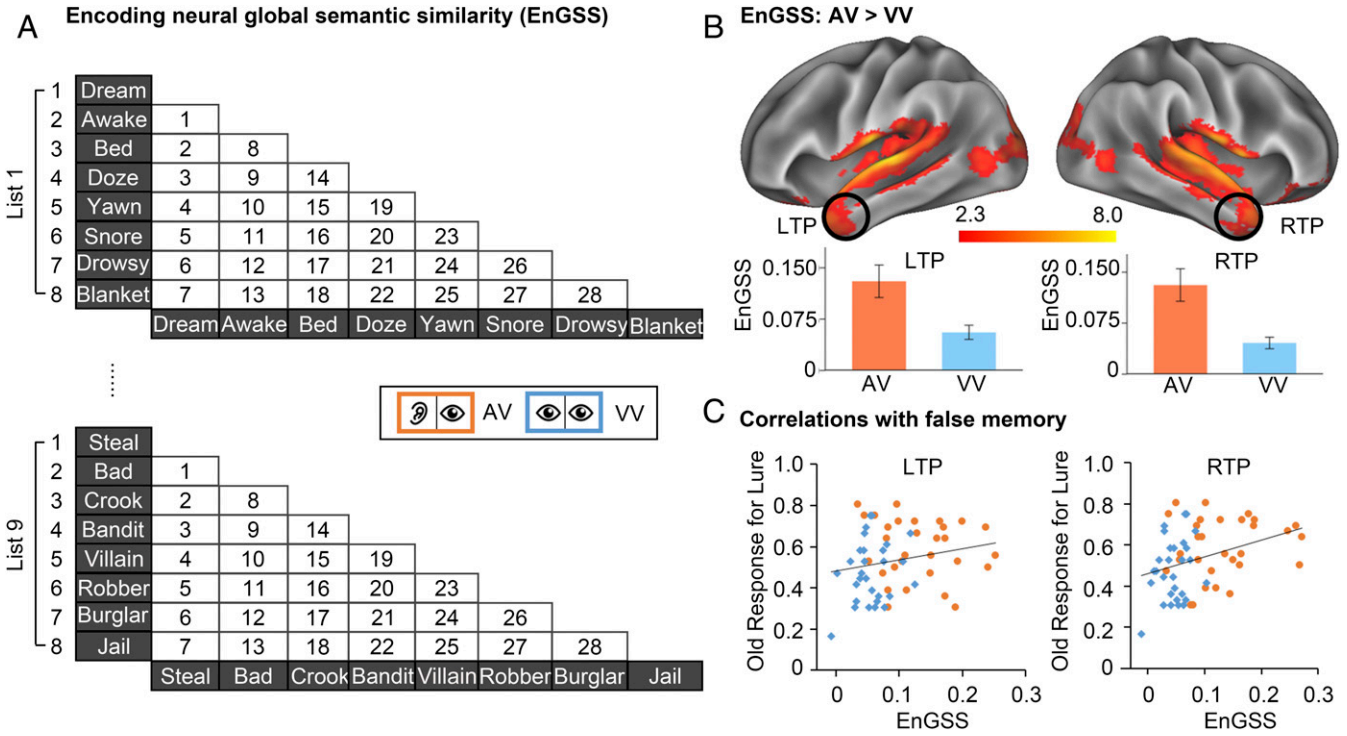

Fig. 4. Greater EnGSS in the AV group than in the VV group. (*A*) The EnGSS was calculated by averaging the Fisher's Z scores reflecting encoding neural activation pattern similarity of 28 pairwise correlations for eight studied items within each word list. It was calculated for the VV and AV groups, separately. (*B*). EnGSS AV > VV: Bilateral temporal pole (LTP and RTP) associated with greater EnGSS in AV than those in VV. Bar graphs of EnGSS in the AV and VV conditions are shown for LTP and RTP. Error bars indicate SEs. (*C*) Greater EnGSS in bilateral temporal pole was associated with higher endorsement rate of lures (judged as old at retrieval). Each dot represents a participant (AV in orange, VV in blue).

strength (TO vs. FN) on the ER-nGPS in these three visual cortex ROIs were still significant ($P < 0.001$).

Discussion

In two experiments we found a consistent modality effect on the rate of false memories. That is, AV was associated with more false memories and fewer true memories than VV. Furthermore, the rate of true memories was affected more by the semantic similarity under the auditory learning conditions than under the visual learning conditions. Using fMRI and representational similarity analysis, we found that compared with the VV group the AV group showed less representational match between encoding and retrieval in the visual cortex, greater semantic encoding in the temporal pole, and weaker prefrontal monitoring during retrieval. These results together provide a deeper mechanistic understanding of the induction of false memory (*SI Appendix*, Fig. S8).

Behavioral results from both experiments confirmed that the AV condition showed higher rates of false memories and lower rates of true memories than did the VV condition. This result was consistent with a meta-analysis of 32 previous DRM experiments, which showed about 10% higher false recognition in AV than VV (29). It should be noted that one previous study found a very different pattern of false memory across conditions (i.e., VA > VV > AV > AA) (30). In that study, researchers used a within-subjects design with only three lures for each modality. This small number of trials might have reduced the stability and reproducibility of the results. In the current study, we used a between-subjects design with 9 word lists and 36 lures (four lures for each list), and our results were consistent with previous experiments using either between-subjects (24 lures per subject) or within-subjects designs (nine lures per modality) (4). Taken together, this behavioral pattern is robust and not affected by the inclusion of a perceptual task between word lists or the experimental conditions under the fMRI setup.

Guided by the TAP and global matching models, we employed a way to link neural representations to true and false memories. In particular, rather than simply examining the encoding–retrieval similarity for a given item, we examined the neural pattern similarity between an item at retrieval and all studied items at encoding (i.e., ER-nGPS). We found that the matching of episodic context between the test item and all studied items in the LMFG was associated with both true and false memories strength in AV and VV. This result replicated and extended a previous finding that the ER-nGPS in the lateral frontal gyrus was associated with both true and false memories (21). In particular, the commonality between auditory and visual modalities suggests that this region contains supramodal mnemonic representation that can be used to guide cross-modal decisions. Consistently, previous studies using visual objects or abstract shapes also found that activation in this region was associated with both true and false memories (31, 32).

We further found several important differences between AV and VV conditions which help to understand how contextual match influences the rates of true and false memories. First, we found higher rates of true memory in VV than AV for both experiments, even though the level of true memory was already quite high in both conditions. This behavioral result replicated previous studies (29) and was consistent with the theoretical framework that emphasizes the match between study and test modalities (4, 11). Consistently, we found greater global matching for true memory in the visual cortex and the LHIP for the VV than AV conditions. The involvement of the visual cortex in true memory under the VV condition replicated the finding of a previous study using MVPA (21). These results suggest that the encoding–retrieval similarity increases true memory (14). Our

finding may also help to explain why the results of visual cortex were mixed in previous studies using univariate analysis. Specifically, early studies using the VV condition found greater visual cortex activation for true memory than false memory (33, 34), whereas those using the AV condition did not (35, 36).

We also found greater ER-nGPS in the hippocampus for the VV than AV conditions, with the LHIP showing true memory signal for VV but not AV. Unlike the visual cortex that mainly reflects the modality match, the hippocampus has been implicated in more general contextual representation (25). In particular, given the dentate gyrus/CA3's role in pattern separation and pattern completion, they should contribute to the encoding–retrieval pattern similarity. Nevertheless, fMRI studies examining the global matching signal in the hippocampus have revealed weak and inconsistent results (21, 23). In both the current study and a previous study (21), no significant hippocampal signal was found using whole-brain searchlight analysis, and only weak differences between TO and FN were found in the LHIP. Two previous studies examining the global matching signal during encoding also reported inconsistent findings (26, 27). One study found that the higher global matching signal in the medial temporal lobe including the hippocampus was associated with true memory (27), whereas the other study using high-resolution fMRI found that lower global matching signal was associated with better true memory (26). Due to the functional heterogeneity of the hippocampus (37) and its sparse representation of item information (38), future studies with high-resolution fMRI (26) and intracranial EEG (39) could help to elucidate the role of hippocampal contextual representation in true and false memories.

Second, we found that semantic coding in the temporal pole (i.e., a "semantic hub" in the brain) also plays a role in how modality influences false memories. Behavioral results suggested that compared with visual learning, mnemonic judgment after auditory learning was more influenced by semantic similarity, likely due to enhanced semantic processing in auditory modality (40). Our fMRI results further suggested that although there were comparable levels of activation between the two modalities (41, 42), auditory learning led to greater representational similarity between semantically related words in the temporal pole, which was in turn correlated with higher false memories. Supporting the role of the temporal pole in semantically mediated true and false memories, previous studies found that transcranial magnetic stimulation (TMS) in the anterior temporal lobe reduced DRM false memory but did not affect true memory (43–45). Our finding was also consistent with a recent fMRI study which revealed that semantic representations in the temporal pole determine the likelihood of false memory (22). Together, the current study provides neural representational evidence that auditory processing of words could enhance semantic-based neural code in the temporal pole, which increases the likelihood of semantic-induced false memory.

Third, we found that compared with the VV condition, the AV condition reduced the involvement of the prefrontal monitoring process, which contributed to the elevated rate of false memories. A strong involvement of the prefrontal cortex in the processing of lures is predicted by the activation/monitoring model (28) and is consistent with several previous studies (46, 47). Intuitively, we would predict that there might be greater need for retrieval search and monitoring in the AV condition, as the context was changed. Nevertheless, the behavioral data indicated there was no difference in reaction time between AV and VV in Exp. 2, and subjects in the AV group were even a little faster than subjects in the VV group in TO and LO in Exp. 1 ($P < 0.05$). It has been recently suggested that the prefrontal cortex's involvement is triggered by the mismatch between different sources of memory signals, including the signals from frontoparietal regions that did not differentiate true and false memories and the signals from the sensory cortex indicating true memory (21). The current study replicated this finding and further suggested that the same mechanism could account for the diminished monitoring process in the AV condition, as both targets and lures rely solely on the frontal memory signals. Consistently, previous behavioral DRM studies suggest that, compared with AV, the distinctive orthographic features from VV may have promoted the monitoring process at retrieval (4, 5, 48). In addition, a previous event-related potential study using common words also revealed a frontal component that could differentiate old and new items in the VV condition but not in the AV condition (49).

The reduced prefrontal monitoring process at retrieval also has significant implications for our understanding of the false memory phenomenon. For example, most existing studies using a misinformation manipulation involved a mismatch in stimulus modality, for example visual presentation of a witnessed event but auditory presentation of misinformation afterward (6–8). As suggested by the current study, this modality mismatch might have reduced the engagement of the monitoring process and increased the chance of successful implantation of false memories. In another study where the misinformation was presented in the same modality, it was found that subsequent true memory was associated with greater activation in the frontoparietal region, whereas subsequent false memories were associated with less involvement of frontoparietal control regions and greater engagement of bilateral temporal cortices (50). Future studies should apply our methods to more ecological experimental designs (8, 51) and directly compare the effect of modality match on implanted false memories. Furthermore, perturbation of prefrontal activity using either transcranial direct-current stimulation and TMS or behavioral manipulation (e.g., a dual task) might help to establish a causal role of the prefrontal cortex in false memories and guide the development of novel brain-based methods for memory manipulations.

Taken together, combining theoretical and computational models and representational similarity analysis of neuroimaging data, the current study demonstrated how a better understanding of the formation of true and false memories under different learning and test circumstances could be achieved by examining the interactions of memory representations during encoding and retrieval. These findings not only help to advance a more mechanistic understanding of our malleable memory systems but also have important practical implications for developing more effective ways to enhance true memory and reduce false memory in our educational, clinical, and legal practices (1).

Methods

Participants. In the behavioral study (Exp. 1), 118 participants (75 females and 43 males, mean age 22.14 ± 2.21 y, ranging from 17 to 28 y) were randomly assigned into four groups (31, 29, 29, and 29 participants in AV, VV, AA, and VA groups, respectively). In the fMRI study (Exp. 2), an independent sample of 59 participants (28 females and 31 males, mean age 21.25 ± 1.45 y, ranging from 18 to 25 y) was randomly assigned into two groups (30 and 29 participants in AV and VV, respectively). Age and gender were carefully matched between groups in both experiments ($P > 0.05$). All participants were right-handed Chinese college students who had normal vision and hearing and no history of psychiatric or neurological diseases. Written consent was obtained from each participant. This study was approved by the Institutional Review Board of the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University.

Materials. Nine word lists, each containing 12 two-character Chinese words that describe one theme, were used in both two studies. They were translated and adapted from materials used in Roediger and McDermott (2). For example, one list included words like "dream," "awake," "bed," "doze," "yawn," "snore," "drowsy," "blanket," "sleep," "rest," "tired," and "pillow." Of the 12 words in each list, 8 words were studied (only four of them would be tested, that is, targets, such as "dream," "bed," "yawn," and "drowsy") and the other four words were used as critical lures (e.g., "sleep,"

"rest," "tired," and "pillow"). In addition, 36 semantically unrelated words (e.g., "visit") were used as foils in the recognition test. Study or test items were presented either visually in the center of the computer screen or auditorily in a female voice.

Experimental Design. There were four experimental conditions according to the stimulus modality during encoding and recognition memory test: auditorily presented during both encoding and test (AA), auditorily presented during encoding and visually presently during test (AV), visually presented during encoding and auditorily presented during test (VA), and visually presented during both encoding and test (VV) (Fig. 1*A*). Experimental materials and design were the same for both experiments, except that Exp. 1 was purely a behavioral study comparing all four conditions, whereas Exp. 2 was an fMRI study that included only the AV and VV conditions.

During the encoding phase, participants were explicitly instructed to intentionally memorize each word presented and were told that there would be a recognition test later. Before studying these words, participants were warned that there would be some unstudied words which were semantically related to the studied words in the following recognition test. As shown in Fig. 1*A*, participants studied 72 words (i.e., nine word lists) over three sessions/runs, each containing three word lists. These words were presented by word list. Before the start of each word list, there was a 1-s visual cue (i.e., "List 1"). Each word was presented only once. The order of these eight studied words within each word list and that of the nine word lists were randomized across participants. For the fMRI study, a slow event-related design (12 s for each trial) was used. Each trial started with a 1-s fixation point, followed by a visually or auditorily presented Chinese word for 1 s. To help the participants to remember these words, participants were asked to make a pleasantness judgment on the word by pressing one of four buttons with their left or right index finger or middle finger (1 = "very unpleasant," 2 = "mildly unpleasant," 3 = "mildly pleasant," and 4 = "very pleasant") within 2 s. Three seconds after the onset of the word, participants were asked to perform a perceptual judgment task for 8 s, which was included to prevent the participants from further processing the studied words. A self-paced procedure was used to make this task more engaging. A Gabor image tilting 45° to the left or the right vertically was randomly presented on the screen, and participants were asked to identify the orientation of the Gabor by pressing two buttons (one left, four right). Participants were asked to respond as quickly and accurately as possible. The next trial started 0.1 s after participants' response.

After studying all nine lists, participants were given a two-back working memory task (using digital numbers as materials) for 10 min before they took the recognition test. The working memory task served as a distractor task between the encoding and retrieval phases and allowed time for an anatomic MRI scan. During the recognition test, participants were asked to judge whether they had studied the visually or auditorily presented words earlier by pressing one of four buttons (1 = "Definitely new," 2 = "Probably new," 3 = "Probably old," and 4 = "Definitely old"). These confidence responses were used to index memory strength. The use of right vs. left hand for old vs. new response was counterbalanced across participants. In total, 108 words (36 target words, 36 critical lures, and 36 foils) were presented over three (scanning) sessions and the order was pseudorandomized. Following the procedure used in previous studies (52, 53), three unstudied foil words were placed at the beginning of each test run. The same slow event-related design (12 s for each trial) as in the study phase was used for the retrieval phase.

Behavioral Analysis. For memory performance, the endorsement rates were calculated for targets (studied words), lures (semantically related unstudied words), and foils (unrelated unstudied words) recognized as old (scored 3 or 4). As was done in a previous study (4), we corrected for the baseline using a high-threshold correction procedure. The corrected true memory was obtained by subtracting false recognition for foils (i.e., foils judged as old, FO) from true recognition for targets (i.e., targets judged as old, TO); similarly, the corrected false memory was obtained by subtracting false recognition for foils (FO) from false recognition for lures (lures judged as old, LO).

For the recognition test in both studies, ANOVA was used to examined the false memory effect (i.e., LO > FO), the effect of modality on the endorsement rates for TO, LO, and FO, on the corrected true and false memories (i.e., TO-FO and LO-FO), and on the reaction time of TO, LO, LN, and FN. To examine the effect of semantic similarity on recognition memory, Pearson correlations were calculated between sGS ratings and the memory scores in the recognition test across items, separately for targets and lures and for each modality group. The sGS was obtained by averaging the semantic similarity of each target or lure to all studied words within a list, which were collected in an independent sample (35 participants) from a previous study using the same materials (21). Thus, each of the 36 targets or each of 36 lures has a specific sGS score. The memory score for each target or lure was obtained by averaging its memory rating (i.e., 1 ~ 4) during the recognition test across participants in the current study, separately for each modality group.

fMRI Data Collection and Preprocessing. All brain imaging scans were performed on a 3.0 T Siemens Magnetom Trio scanner at Beijing Normal University Brain Imaging Center. A single-shot T2*-weighted gradient-echo, echo-planar imaging (EPI) sequence was used for functional imaging acquisition with the following parameters: TR/TE/θ = 2,000 ms/25 ms/90°, field of view (FOV) = 192 × 192 mm, matrix = 64 × 64, and slice thickness 3.0 mm. Forty-one contiguous axial slices parallel to the AC–PC line were obtained to cover the whole cerebrum and partial cerebellum.

Structural MRI was acquired using a T1-weighted, 3D, gradient-echo pulse-sequence (MPRAGE). For subjects in the AV group, the parameters for this sequence were T1/TR/TE/θ = 1,100 ms/2,530 ms/3.39 ms/7°, FOV = 256 × 256 mm, matrix = 256 × 256, and slice thickness = 1.33 mm. A total of 144 sagittal slices were acquired to provide high-resolution structural images of the whole brain. For subjects in the VV group, a refined structural MRI was acquired, and parameters for this sequence were T1/TR/TE/θ = 800 ms/2,530 ms/3.09 ms/10°, FOV = 256 × 256 mm, matrix = 256 × 256, and slice thickness = 1 mm. A total of 208 sagittal slices were acquired to provide high-resolution structural images of the whole brain.

The FEAT (FMRI Expert Analysis Tool) version 6.00, part of the FSL (FMRIB software library, version 5.0.9; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki) was used for image preprocessing and statistical analysis. To allow for T1 equilibrium, the scanner discarded the first three volumes before the task automatically. Then, the remaining images were realigned to correct for head movements. For all participants and sessions, translational movement parameters never exceeded one voxel in any direction. A 5-mm FWHM Gaussian kernel was used for spatially smoothing of the data. The data were then filtered temporally using a nonlinear high-pass filter with a 90-s cutoff. We used a two-step registration procedure (e.g., EPI images were first registered to the MPRAGE structural image and then into the standard MNI space using affine transformations). The FNIRT nonlinear registration was used for further refining the registration from structural images to the standard space. Models were constructed using the general linear model within the FILM module of FSL.

Single-Item Response Estimation. We used general linear model to compute the beta map for each of the 72 studied words during encoding and 108 words during retrieval in the VV and AV groups, separately (Fig. 2*A*). The ER-nGPS was calculated by averaging Fisher's Z scores of neural activation pattern similarity (Pearson *r*) between each item (i) during retrieval (R) with all other items (j) during encoding (E, from 1 to n). In this single-trial model, the presentation of each stimulus was modeled as an impulse, and it convolved with a conical hemodynamic response function (double gamma) (54). To obtain reliable estimates of single trial responses, the lease-square single method was used. The beta values of each stimulus were used to calculate the neural pattern similarity and used in the following statistical analysis.

Neural Global Pattern Similarity Between Encoding and Retrieval (ER-nGPS). Using the searchlight method, we identified the brain regions whose neural global pattern similarity between encoding and retrieval was associated with memory strength (55). For each voxel, we extracted signals from the cubic ROI containing 125 surrounding voxels. For each tested item, we calculated pairwise Pearson correlations between the activation patterns of this item during retrieval with the activation pattern of all studied words during encoding. We transformed these similarity scores into Fisher's Z scores and then averaged them to generate the ER-nGPS value. Four types of trials were modeled, including targets judged as old (TO), lures judged as old (LO), lures judged as new (LN), and foils judged as new (FN). The foils judged as old and targets judged as new were not included because they were rare. The ER-nGPS for each types of trials (TO, LO, LN, and FN) were then separately averaged and contrasted at individual level. A random-effects model was used for the group analysis. Since no first-level variance was available, an ordinary least squares model was used.

First, we examined whether the ER-nGPS could predict both true and false memories in AV or VV. Next, using a conjunction analysis (56) with the easythresh.conj script in FSL, we explored the ER-nGPS for both AV and VV groups, by examining the contrast (TO minus FN) shared by the AV and VV groups. Moreover, we explored ER-nGPS differences (TO minus FN) on each of the contrasts (AV minus VV or VV minus AV). Because we were interested

in these neural indices in brain regions showing the effect of memory strength in the AV and VV groups, we used a relatively liberal threshold to find these regions ($Z > 2.3$ and a cluster probability of $P < 0.05$, corrected for whole-brain multiple comparison using Gaussian random filed theory). Unless otherwise noted, the same threshold was used for all of ER-nGPS, EnGSS, and univariate analyses. These analyses have sufficient power using the current sample size based on previous studies (21, 57, 58).

Regions showing true memory effect for both VV and AV (the LMFG and LING) and for VV only (the LdLOC, LvLOC, and RvLOC) were defined as ROIs. These five ROIs were defined by including all of the voxels in each cluster showing suprathreshold activation for each contrast. The mean ER-nGPS and univariate activations of these ROIs for TO, LO, LN, and FN in AV and VV were then extracted and analyzed. In these five ROIs, ANOVA was used to examine whether the ER-nGPS was associated with false memories strength (i.e., LO > LN) in AV and VV. We also examined the effect of ER-nGPS on true and false memories after controlling for univarite activations in each of these ROIs.

In addition, due to the special role of hippocampus in memory, we further examined the ER-nGPS in the anatomically defined hippocampus ROIs, based on the Harvard–Oxford probabilities atlas (*SI Appendix*, Fig. S4).

EnGSS. We used MVPA to examine the EnGSS between eight studied words within each word list for the AV and VV groups, separately (Fig. 4*A*). The searchlight method ($5 \times 5 \times 5$ voxels cubic) was used. For a searchlight sphere, the multivoxel response pattern for each of the 28 pairs of correlations between these eight studied words within each list was extracted for each participant separately. All similarity scores were transformed into Fisher's Z scores for further statistical analysis. Pattern similarities were estimated by calculating the pairwise Pearson correlation among each trial's response pattern within each word list and then averaged across nine word lists. To examine the modality-specific encoding, we compared different modality (AV minus VV or VV minus AV). The searchlight analysis was conducted in the native space for each participant and then transformed into standard space for group analysis. A random-effects model was used for group analysis. Because no first-level variance was available, an ordinary least squares model was used.

Because the temporal pole plays a key role in semantic encoding related to false memory (22), we defined two ROIs (the LTP and RTP) by including all voxels showing significantly greater EnGSS for AV than VV within the anatomical boundary of temporal pole, based on the Harvard–Oxford template. The mean EnGSS of these ROIs in AV and VV were then extracted and correlated with the TO and LO across subjects. Next, across nine word lists in the AV or VV group, we explored the correlations between the EnGSS in TP and the mean false memory rate (averaged across four critical lures for each word list).

Univariate Activation-Based Analysis for the Retrieval Phase. During the retrieval phase, four types of trials were modeled, including TO, LO, LN, and FN. We explored the univariate activations for true memory (TO vs. FN), false memory (LO vs. LN), and monitoring process (Lure [LO and LN] vs. FN) in AV and VV, separately. A higher-level analysis involved cross-run contrasts for each participant used a fixed-effects model. These contrasts were then used for group analysis with a random-effects model, using full FMRIB's Local Analysis of Mixed Effect 1+2 with automatic outlier detection (59, 60). We examined the common activation for AV and VV using a conjunction analysis (56). To examine our hypothesis, we further examined the activation differences between AV and VV in monitoring processes, using the small volume correction to restrict our search in the left inferior and middle frontal gyri based on previous studies (21). Focusing on the LPFC region showing different activations for AV and VV in monitoring processing (by including all voxels showing suprathrehold activation), we correlated the univariate activation for lures in the LPFC and the ER-nGPS difference (LMFG – LvLOC) (i.e., calculated by subtracting the ER-nGPS in the LvLOC from that in the LMFG), separately for VV and AV groups.

Data Availability. Data and materials are available at https://openneuro.org/datasets/ds001650.

ACKNOWLEDGMENTS. We thank Prof. Craig Stark for helpful comments on an early version of the manuscript. This study was supported by National Natural Science Foundation of China Grants 31730038 and 31571132, 973 Program Grant 2014CB846102, National Natural Science Foundation of China (NSFC) and DFG joint project NSFC 61621136008/DFG TRR-169, and the Guangdong Pearl River Talents Plan Innovative and Entrepreneurial Team Grant 2016ZT06S220.

1. Loftus E (2003) Our changeable memories: Legal and practical implications. *Nat Rev Neurosci* 4:231–234.
2. Roediger HL, McDermott KB (1995) Creating false memories: Remembering words not presented in lists. *J Exp Psychol Learn Mem Cogn* 21:803–814.
3. Smith RE, Hunt RR (1998) Presentation modality affects false memory. *Psychon Bull Rev* 5:710–715.
4. Gallo DA, McDermott KB, Percer JM, Roediger HL, 3rd (2001) Modality effects in false recall and false recognition. *J Exp Psychol Learn Mem Cogn* 27:339–353.
5. Pierce BH, Gallo DA, Weiss JA, Schacter DL (2005) The modality effect in false recognition: Evidence for test-based monitoring. *Mem Cognit* 33:1407–1413.
6. Zhu B, et al. (2010) Individual differences in false memory from misinformation: Cognitive factors. *Memory* 18:543–555.
7. Loftus EF (2005) Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learn Mem* 12:361–366.
8. Stark CE, Okado Y, Loftus EF (2010) Imaging the reconstruction of true and false memories using sensory reactivation and the misinformation paradigms. *Learn Mem* 17:485–488.
9. Xue G (2018) The neural representations underlying human episodic memory. *Trends Cogn Sci* 22:544–561.
10. Morris CD, Bransford JD, Franks JJ (1977) Levels of processing versus transfer appropriate processing. *J Verbal Learn Verbal Behav* 16:519–533.
11. Tulving E, Thomson DM (1973) Encoding specificity and retrieval processes in episodic memory. *Psychol Rev* 80:352–373.
12. Roediger HL, Guynn MJ (1996) Retrieval processes. *Memory*, eds Bjork EL, Bjork RA (Academic, New York), pp 197–236.
13. Godden DR, Baddeley AD (1975) Context-dependent memory in two natural environments: On land and underwater. *Br J Psychol* 66:325–331.
14. Ritchey M, Wing EA, LaBar KS, Cabeza R (2013) Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cereb Cortex* 23:2818–2828.
15. Clark SE, Gronlund SD (1996) Global matching models of recognition memory: How the models match the data. *Psychon Bull Rev* 3:37–60.
16. Humphreys MS, Pike R, Bain JD, Tehan G (1989) Global matching: A comparison of the SAM, Minerva II, matrix, and TODAM models. *J Math Psychol* 33:36–67.
17. Murdock BB (1982) A theory for the storage and retrieval of item and associative information. *Psychol Rev* 89:609–626.
18. Gillund G, Shiffrin RM (1984) A retrieval model for both recognition and recall. *Psychol Rev* 91:1–67.
19. Hintzman DL (1984) MINERVA 2: A simulation model of human memory. *Behav Res Methods Instrum Comput* 16:96–101.
20. Pike R (1984) Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychol Rev* 91:281–294.
21. Ye Z, et al. (2016) Neural global pattern similarity underlies true and false memories. *J Neurosci* 36:6792–6802.
22. Chadwick MJ, et al. (2016) Semantic representations in the temporal pole predict false memories. *Proc Natl Acad Sci USA* 113:10180–10185.
23. Kurkela KA, Dennis NA (2016) Event-related fMRI studies of false memory: An activation likelihood estimation meta-analysis. *Neuropsychologia* 81:149–167.
24. Schacter DL, Loftus EF (2013) Memory and law: What can cognitive neuroscience contribute? *Nat Neurosci* 16:119–123.
25. Davachi L, DuBrow S (2015) How the hippocampus preserves order: The role of prediction and context. *Trends Cogn Sci* 19:92–99.
26. LaRocque KF, et al. (2013) Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *J Neurosci* 33:5466–5474.
27. Davis T, Xue G, Love BC, Preston AR, Poldrack RA (2014) Global neural pattern similarity as a common basis for categorization and recognition memory. *J Neurosci* 34:7472–7484.
28. Roediger HL, 3rd, Watson JM, McDermott KB, Gallo DA (2001) Factors that determine false recall: A multiple regression analysis. *Psychon Bull Rev* 8:385–407.
29. Gallo DA (2006) *Associative Illusions of Memory: False Memory Research in DRM and Related Tasks* (Psychology, London), pp 114–116.
30. Maylor EA, Mo A (1999) Effects of study-test modality on false recognition. *Br J Psychol* 90:477–493.
31. Garoff RJ, Slotnick SD, Schacter DL (2005) The neural origins of specific and general memory: The role of the fusiform cortex. *Neuropsychologia* 43:847–859.
32. Garoff-Eaton RJ, Slotnick SD, Schacter DL (2006) Not all false memories are created equal: The neural basis of false recognition. *Cereb Cortex* 16:1645–1652.
33. Kim H, Cabeza R (2007) Differential contributions of prefrontal, medial temporal, and sensory-perceptual regions to true and false memory formation. *Cereb Cortex* 17:2143–2150.
34. Dennis NA, Kim H, Cabeza R (2007) Effects of aging on true and false memory formation: An fMRI study. *Neuropsychologia* 45:3157–3166.
35. Abe N, et al. (2008) Neural correlates of true memory, false memory, and deception. *Cereb Cortex* 18:2811–2819.
36. Cabeza R, Rao SM, Wagner AD, Mayer AR, Schacter DL (2001) Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. *Proc Natl Acad Sci USA* 98:4805–4810.
37. Bakker A, Kirwan CB, Miller M, Stark CE (2008) Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science* 319:1640–1642.
38. Quiroga RQ, Kreiman G, Koch C, Fried I (2008) Sparse but not 'grandmother-cell' coding in the medial temporal lobe. *Trends Cogn Sci* 12:87–91.

39. Lohnas LJ, et al. (2018) Time-resolved neural reinstatement and pattern separation during memory decisions in human hippocampus. *Proc Natl Acad Sci USA* 115: E7418–E7427.
40. Holcomb PJ, Neville HJ (1990) Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Lang Cogn Process* 5:281–312.
41. Booth JR, et al. (2002) Modality independence of word comprehension. *Hum Brain Mapp* 16:251–261.
42. Chee MW, O'Craven KM, Bergida R, Rosen BR, Savoy RL (1999) Auditory and visual word processing studied with fMRI. *Hum Brain Mapp* 7:15–28.
43. Boggio PS, et al. (2009) Temporal lobe cortical electrical stimulation during the encoding and retrieval phase reduces false memories. *PLoS One* 4:e4959.
44. Díez E, Gómez-Ariza CJ, Díez-Álamo AM, Alonso MA, Fernandez A (2017) The processing of semantic relatedness in the brain: Evidence from associative and categorical false recognition effects following transcranial direct current stimulation of the left anterior temporal lobe. *Cortex* 93:133–145.
45. Gallate J, Chi R, Ellwood S, Snyder A (2009) Reducing false memories by magnetic pulse stimulation. *Neurosci Lett* 449:151–154.
46. Garoff-Eaton RJ, Kensinger EA, Schacter DL (2007) The neural correlates of conceptual and perceptual false recognition. *Learn Mem* 14:684–692.
47. Kim H, Cabeza R (2007) Trusting our memories: Dissociating the neural correlates of confidence in veridical versus illusory memories. *J Neurosci* 27:12190–12197.
48. Pierce BH, Gallo DA (2011) Encoding modality can affect memory accuracy via retrieval orientation. *J Exp Psychol Learn Mem Cogn* 37:516–521.
49. Curran T, Dien J (2003) Differentiating amodal familiarity from modality-specific memory processes: An ERP study. *Psychophysiology* 40:979–988.
50. St Jacques PL, Olm C, Schacter DL (2013) Neural mechanisms of reactivation-induced updating that enhance and distort memory. *Proc Natl Acad Sci USA* 110:19671–19678.
51. Okado Y, Stark CE (2005) Neural activity during encoding predicts false memories created by misinformation. *Learn Mem* 12:3–11.
52. Duverne S, Motamedinia S, Rugg MD (2009) The relationship between aging, performance, and the neural correlates of successful memory encoding. *Cereb Cortex* 19:733–744.
53. Gerard L, Zacks RT, Hasher L, Radvansky GA (1991) Age deficits in retrieval: The fan effect. *J Gerontol* 46:P131–P136.
54. Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage* 59:2636–2643.
55. Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci USA* 103:3863–3868.
56. Nichols T, Brett M, Andersson J, Wager T, Poline JB (2005) Valid conjunction inference with the minimum statistic. *Neuroimage* 25:653–660.
57. Mumford JA, Davis T, Poldrack RA (2014) The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *Neuroimage* 103:130–138.
58. Desmond JE, Glover GH (2002) Estimating sample size in functional MRI (fMRI) neuroimaging studies: Statistical power analyses. *J Neurosci Methods* 118:115–128.
59. Woolrich MW, Behrens TE, Beckmann CF, Jenkinson M, Smith SM (2004) Multilevel linear modelling for FMRI group analysis using Bayesian inference. *Neuroimage* 21: 1732–1747.
60. Beckmann CF, Jenkinson M, Smith SM (2003) General multilevel linear modeling for group analysis in FMRI. *Neuroimage* 20:1052–1063.
61. Marcus DS, et al. (2011) Informatics and data mining tools and strategies for the human connectome project. *Front Neuroinform* 5:4.